

# Contraction of fully expanded FMR1 alleles to the normal range: predisposing haplotype or rare events?

**Nuno Maia, Joana R Loureiro, Bárbara Oliveira, Isabel Marques, Rosário Santos, Paula Jorge and Sandra Martins**

Originally published in the *Journal of Human Genetics* 2017 Feb;62(2):269-275.

## ABSTRACT

Fragile X syndrome (FXS), the most common cause of inherited intellectual disability, is due to the expansion over 200 CGGs and methylation of this polymorphic region, in the 5' -UTR (untranslated region) of FMR1 (Xq27.3). We have identified four FXS mosaic males: M1-(CGG)<sub>35</sub>/(CGG)<sub>4200</sub>; M2-(CGG)<sub>26</sub>/(CGG)<sub>4200</sub>; M3-(CGG)<sub>39</sub>/(CGG)<sub>4200</sub>; and M4-(CGG)<sub>18</sub>/(CGG)<sub>125</sub>/(CGG)<sub>4200</sub>. After genotyping their respective mothers, we suggested that normal alleles of these patients resulted from post-zygotic contractions of full expansions. The detection of these four rare independent cases led us to hypothesize the existence of a large-contraction predisposing haplotype in our population. Next, we questioned whether other normal pure CGGs would have arisen through similar contractions from fully expanded alleles. To address these questions, we identified stable single-nucleotide polymorphism (SNP) lineages and related short tandem repeat (STR) haplotypes (DXS998-DXS548-FRAXAC1-FRAXAC2) of the four mosaics, 123 unrelated FXS patients and 212 controls. An extended flanking haplotype (34-44-38-336) shared by mosaics from lineage A suggested a risk lineage-specific haplotype more prone to large contractions. Other normal pure FMR1 alleles from this SNP background also shared phylogenetically close STR haplotypes, although a single (CGG)<sub>exp4</sub>(CGG)<sub>24</sub> contraction or the loss of AGG interruptions may explain their origin. In both scenarios, multistep FMR1 mutations involving the gain or loss of several CGGs seem to underlie the evolution of the repeat.

## INTRODUCTION

Mutations in the Fragile X Mental Retardation 1 (FMR1) gene, encoding the FMR1 Protein (FMRP), are the leading heritable form of intellectual disability and the main monogenic cause of autism. FMR1 contains a 5' noncoding CGG repeat region that may expand beyond the normal range and cause fragile X syndrome (FXS) if fully mutated (4200 CGGs) or other fraxopathies (fragile X tremor/ataxia syndrome, FXTAS and fragile X premature ovarian failure, FXPOF) if in the premutation range (550CGGso200). Full mutations are generally methylated in the promoter region, with the consequent transcriptional silencing and absence of FMRP translation.<sup>1</sup> On the basis of repeat size and instability, FMR1 alleles can be designated as normal (stable 5–44 CGGs), intermediate (slightly unstable 45–54 CGGs) and premutated (very unstable 55–200 CGGs, with the risk of expansion to full mutation dependent on the progenitor gender and allele size).<sup>2–4</sup> A complex mechanism involving multiple steps seems to be on the origin of de novo FMR1 expansions, reason why this has been under debate for the last decades.<sup>5–8</sup> Cis-acting factors such as AGG interspersions in the CGG repeat region are, however, well known important stability elements.<sup>9–12</sup> Normal alleles have AGG interruptions usually after every 9 or 10 CGG triplets,<sup>13</sup> whereas in other allele classes, the number of AGGs tends to be progressively lower as the repeat size increases. In approximately 50% of premutated alleles, an AGG interruption is observed, reducing the risk of expansion upon maternal transmission, particularly in alleles with less than 100 CGGs.<sup>11–16</sup> The loss of AGGs occurs in a polarized way at the 3' - end, creating a long pure (CGG)<sub>n</sub> with higher mutability.<sup>17,18</sup> On the other hand, while analyzing the interspersions pattern of the repeat, an increased instability seems to be correlated with the presence of the first interruption at the tenth triplet, as the configuration (CGG)<sub>9</sub>AGG (CGG)<sub>n</sub> is frequently observed in large normal alleles.<sup>19</sup> In addition, normal-sized alleles with more than 24 pure CGGs at the 3' -end frequently share the typically studied DXS<sub>548</sub>-FRAXAC<sub>1</sub> FXS haplotype, suggesting that these alleles could be more prone to expand.<sup>15,17,20,21</sup> FMR1 mosaicism has been reported, with multiple CGG repeat sizes (normal and full mutation alleles) present in different tissues (inter-tissue mosaic) or in different cells from the same tissue (intra-tissue mosaic) of a single patient. These post-zygotic events, utterly responsible for the intergenerational changes on repeat sizes when unstable alleles are further transmitted, may help to elucidate mitotic instability processes. Although rare, repeat size mosaicism has been previously described in FXS patients, some of them males who harbor two or more differently sized expansions and, in a few rare cases, expanded and normal alleles.<sup>22,23</sup> During FXS diagnostic routine, we identified four mosaic males with a typical fragile X phenotype, who carried both a full mutation and a normal pure allele at FMR1. After observing that an extremely rare event of contraction occurred independently in all four patients, we hypothesized the presence of a predisposing haplotype to large contractions in our population. We started by identifying stable single-nucleotide polymorphism (SNP) lineages of both normal pure and expanded alleles. If a common ancestry underlies these two groups of alleles, one should expect SNP lineages to be shared between them. A further comparison of fast-evolving markers (STRs) within each lineage allowed us to reconstruct phylogenetic networks, crucial to visualize genetic distances among FMR1 alleles and, thus, to clarify the origin of normal pure alleles.

## Subjects and Methods

### Subjects

In the course of the FMR1 analysis routine using previously published methods,<sup>24</sup> four mosaic cases (M1–M4) were identified in the blood samples from male patients with FXS. In addition, we analyzed 212 males with a FMR1 allele within a CGG normal–intermediate range (cohort 1) and 123 unrelated male individuals with a clinical diagnosis of FXS and an expansion over 200 CGGs (cohort 2). The peripheral blood samples were collected after obtaining informed consent from all the subjects enrolled in this study or their legal representatives. Genomic DNA was extracted by standard salting-out procedures. Centro Hospitalar do Porto (CHP, E.P.E.) scientific and ethical committees gave their approval for this project (154-DEFI/193-CES). The proband and four other individuals from a Nigerian family have been genotyped in this study after giving their consent. We performed the identification of MJD lineages<sup>7,12</sup> and genotyping of allelic variants along a 4 kb region flanking the MJD expansion<sup>8</sup> as previously described.

### (CGG)<sub>n</sub> genotyping and tract configuration

The AGG interspersed pattern of FMR1 alleles was assessed by triplet-primed PCR using 150 ng of genomic DNA, 1X PCR Master Mix (Promega, Madison, WI, USA), 5.1 mM of MgCl<sub>2</sub>, 6.4% of dimethyl sulfoxide, 2.6 mM of Betaine, 10 μM primers 1 and 2 and 7.5 μM of primer 3. The PCR reactions were performed using the following cycle conditions: an initial denaturation of 10 min, 45 cycles of 96 °C for 45 s, 58 °C for 45 s, 68 °C for 2 min and a final extension of 10 min for 68 °C. After amplification fragments were separated by capillary electrophoresis in a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and AGG pattern was analyzed using GeneMapper v4.0 software (Applied Biosystems).

### Genotyping of flanking polymorphic markers

Three SNPs from intronic FMR1 regions were selected: rs971000 (intron 1), rs29282 (intron 7) and rs25715 (intron 9). The SNPs were genotyped by high-resolution melting curve analysis in Rotorgene 6000 (Qiagen, Hilden, Germany). The PCR reactions were carried out using 1 × Master Mix, 1 × Evagreen (Biotium, Hayward, CA, USA), 5 μM of each primer and 7.5 ng of gDNA and a PCR program containing an initial denaturation temperature of 95 °C for 5 min, followed by 35 cycles of 95 °C for 45 s, 30 s of annealing temperature, 72 °C for 30 s and a final extension of 10 min at 72 °C (Supplementary Table S1). The high-resolution melting curve analysis was performed using standard parameters and according to the manufacturer's instructions.

The selected STR markers DXS998, DXS548, FRAXAC1 and FRAXAC2 are embedded in three distinct haplotypic blocks, with three of them upstream and FRAXAC2 located downstream the FMR1 repetitive region. Allele nomenclature follows the one previously proposed.<sup>25</sup> STRs were amplified in a multiplex PCR using Multiplex PCR Master Mix (Qiagen) and primers labeled with different fluorochromes. The reactions were done with 1X Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 150 ng of genomic DNA and a primer mix containing 4.5 μM of each DXS998 primer, 8 μM of each DXS548 primer and 10 μM of each FRAXAC1 and FRAXAC2 primer. The PCRs were performed using an initial denaturation for 15 min at 95 °C, 45 cycle steps of 30 s at 95 °C, 1 min and 30 s at 58 °C and 1 min and 30 s at 72 °C, following a final extension at 72 °C during 10 min

(Supplementary Table S1). The PCR products were analyzed by capillary electrophoresis in ABI3130xl and fragment sizes determined using GeneMapper v2.6.0 software. To assess the repeat configuration of the complex microsatellite FRAXAC2, (GT)xC(TA)yTz, PCR reactions contained 100 ng of genomic DNA, 1X PCR Master Mix (Promega, Madison, WI, USA) and 10 pmol of each primer. Amplification started with an initial denaturing step at 95 °C for 10 min, followed by 40 cycles (95 °C for 1 min; annealing temperature for 1 min; and 72 °C for 2 min) and a final extension step at 72 °C during 10 min. The PCR products were purified using the Illustra ExoStar 1-Step (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK) according to the manufacturer's instructions, and sequenced using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). The sequencing reaction consisted of an initial denaturing step at 96 °C for 1 min, followed by 27 cycles (96 °C for 10 s; 50 °C for 5 s; and 60° for 1 min and 15 s), ending with extension at 60 °C for 5 min. The asymmetric PCR fragments were electrophoresed on an ABI PRISM 3130xl Genetic Analyzer and analyzed using the SeqScape v2.5 software (Applied Biosystems).

### Haplotype analysis

Taking advantage of the location on chromosome X of our locus of interest, FMR1, haplotypes were directly assessed on male samples. Phylogenetic relationships among normal (cohort 1) and FXS (cohort 2) alleles were investigated through reduced median followed by median-joining networks, calculated to reduce some of the reticulation of analyzed microsatellite markers (Network 4.613 software; [www.fluxus-engineering.com](http://www.fluxus-engineering.com)).

## Results

We identified four very rare cases showing intra-tissue mosaicism: FXS male patients harboring, in addition to the full mutation, a normal FMR1 allele (patient M4 carried an additional premutation of 125 CGGs). To gain insight into the mechanism of instability underlying the origin of these alleles, we genotyped the maternal FMR1 repeats transmitted to these mosaic patients (Table 1). Normal alleles carried by mothers differed from those found in patients regarding the number of CGGs. This suggested that post-zygote large contractions from expanded alleles occurred in our diagnosed patients. Taking into account that such a rare event occurred independently in several patients, we hypothesized the existence of an FMR1 haplotype prone to large contractions. With these contractions occurring either recurrently in the same genetic background or in different haplotypes, we wanted next to investigate the origin of all normal-sized alleles in our control population. If some fully expanded CGGs are able to contract to the normal range, the pool of normal alleles may include not only those evolving at a regular rate over the course of many thousands of years, but also others resulted, more recently, from large contractions. To follow the evolutionary history of highly mutable loci, such as the CGG repeat at FMR1, one should start by looking at a narrower time frame. As events on the origin of SNP are unique during the evolution of a species, we defined very stable FMR1 backgrounds by assessing SNP haplotypes (lineages) of this region. Next, the analysis of fast-evolving loci, such as STRs, allowed us to follow a more recent mutational history of the FMR1 alleles.

### SNP–STR haplotypes of mosaic FXS patients

Through the analysis of rs971000, rs29282 and rs25715 (Supplementary Table S2), we assessed SNP backgrounds of our mosaic patients and found two different haplotypes: A-T-T (lineage A) in M1 and M2 and G-C-C (lineage C) in M3 and M4. Next, the genotyping of flanking STRs (DXS998-DXS548-FRAXAC1-FRAXAC2) has shown an extended haplotype shared by mosaics from lineage A, A21: 34-44-38-336 (Table 2). The fact that two independent large contractions (observed in mosaics M1 and M2) have occurred on the same SNP-STR background led us to question the existence of a lineage-specific predisposing haplotype. Interestingly, mosaics M3 and M4 from lineage C also shared a FRAXAC1-FRAXAC2 haplotype: 36-333. Taking into account that upstream markers DXS998 and DXS548 are located in more distant haplotypic blocks, a recombination event might have occurred during the meiosis of maternal FMR1 alleles, being responsible for breaking up a longer haplotype common to mosaics from lineage C.

Regarding the complex polymorphic structure of FRAXAC2 (GT) $\times$ C(TA) $\gamma$ T $\gamma$ , one could argue that same-sized FRAXAC2 alleles had different configuration patterns, thus, not sharing identity-by-state. To discard this possibility, we sequenced all four FRAXAC2 alleles from mosaic cases: alleles 336 from M1 and M2 shared the (GT) $_{16}$ C(TA) $_{7}$ T $_{13}$  configuration, similarly to alleles 333 from M3 and M4, which shared the (GT) $_{14}$ C(TA) $_{7}$ T $_{14}$  pattern.

### Repeat size and configuration of FMR1 alleles on the identified SNP lineages

The identification of FMR1 tract configurations from our mosaic patients has shown all normal alleles carrying no AGG interruptions, henceforward called pure alleles. We then assessed the repeat size and configuration of all the 212 healthy controls (cohort 1) to next analyze the origin of pure CGGs found in the pool of normal FMR1 alleles. After having evidence of four normal pure CGGs originated by four independent large contractions of expanded FMR1 alleles, we aimed at testing whether other uninterrupted repeats from our control population have (1) resulted by the same process or (2) evolved from other normal alleles by losing AGG interruption(s). By analyzing the same three SNPs as in the mosaic cases, we were able to place all FMR1 alleles on stable lineages. In cohort 1, normal and intermediate alleles ranged between 8 and 53 repeat units. The most common pattern was (CGG) $_{10}$ AGG(CGG) $_{9}$ AGG(CGG) $_{9}$  (rf=0.344), followed by (CGG) $_{9}$ AGG(CGG) $_{9}$ AGG(CGG) $_{9}$  (rf=0.127) and (CGG) $_{10}$ AGG(CGG) $_{9}$  (rf=0.090; Supplementary Table S3). The first AGG occurs more frequently after 10 (rf=0.594), 9 (rf=0.283) or 13 (rf=0.047) CGGs, similar to other worldwide populations.<sup>13</sup> Interestingly, we have found eight normal pure alleles distributed in lineages: A-T-T (n=5), G-T-C (n=2) and G-C-C (n=1; Table 3). Under mere stochasticity, the observed proportions of each line (62.5, 25 and 12.5%, respectively) would favor the origin of these normal pure alleles from the pool of normal interspersed repeats, as A-T-T (60%), G-T-C (25.5%) and G-C-C (7.3%) are the most frequent lineages in a total of 55 control alleles with one AGG interruption (Supplementary Table S4). On the contrary, among all analyzed expanded alleles, A-T-T accounts for only 10.6% (G-T-C: 59.3%; G-C-C: 22.8%; Table 3). Taking into account, however, that risk factors may predispose any haplotypic background to instability (independent of its frequency), both hypotheses urged to be tested.

### Phylogenetic distances between normal and fully expanded alleles

We further analyzed the STR backgrounds of all normal pure FMR1 alleles genotyped in cohort 1, as mosaic patients from the same SNP lineage shared entirely or partially a DXS998-DXS548-FRAXAC1-FRAXAC2 haplotype: 34-44-38-336 in M1 and M2 (lineage A) and a 3'-end 36-333 in M3

and M<sub>4</sub> (lineage C; Table 2). All five pure CGGs from lineage A shared the FRAXAC<sub>1</sub> allele, with DXS<sub>998</sub> and FRAXAC<sub>2</sub> alleles also shared by all except one individual. A similar scenario was observed for lineage C even if a single pure CGG has been genotyped: the 32-40-36-333 haplotype (segregated with (CGG)<sub>17</sub>) is only a stepwise mutation apart from the haplotype identified in case M<sub>3</sub> (32-42-36-333). Again, by sequencing, we confirmed that FRAXAC<sub>2</sub> alleles from pure FMR<sub>1</sub> repeats of lineage A (336) were identical by state to the ones genotyped in M<sub>1</sub> and M<sub>2</sub>: (GT)<sub>16</sub>C(TA)<sub>7</sub>T<sub>13</sub>. In addition, allele 338 segregating with (CGG)<sub>31</sub> is only a step mutation apart from all the others: (GT)<sub>17</sub>C(TA)<sub>7</sub>T<sub>13</sub>. Likewise, in lineage C, allele 333 shares the same configuration as same-sized FRAXAC<sub>2</sub> alleles from M<sub>3</sub> and M<sub>4</sub>.

The fact that normal pure FMR<sub>1</sub> alleles from our control population are phylogenetically close to normal alleles from mosaic patients (resulted from large contractions) supported the hypothesis of their origin by similar large contraction events from fully expanded FMR<sub>1</sub> alleles, on a predisposing haplotype. On the other hand, if by chance, this STR haplotype was frequent among interrupted FMR<sub>1</sub> alleles, the alternative hypothesis for the origin of pure CGG alleles by the loss of AGGs would be equally likely. Therefore, we studied 123 additional patients with a full methylated expansion over 200 CGGs (cohort 2) to compare phylogenetic relationships between normal pure alleles from the control population and (1) fully expanded alleles (Figure 1a) or (2) normal alleles carrying one AGG interruption (Figure 1b). Given the scarcity of G-T-C and G-C-C normal pure alleles, we proceeded with next analyses exclusively in lineage A (A-T-T). All normal pure haplotypes were phylogenetically close to each other, with a maximum of only three mutations apart. While testing the first hypothesis, they are shown clustered in a branch of the network, with a single haplotype (A<sub>2</sub>) shared between normal pure and FXS alleles. This suggests that recurrent large contractions from fully expanded alleles do not explain the origin of the five normal pure A-T-T alleles genotyped in cohort 1. Instead, if a single contraction occurred, a (CGG)<sub>24</sub> allele would have arisen in haplotype A<sub>2</sub>, later evolving to A<sub>1</sub> (after a stepwise mutation in DXS<sub>998</sub>) and to A<sub>4</sub> (after stepwise mutations in DXS<sub>548</sub> and in the CGG repeat). Further multistep mutations in the (CGG)<sub>n</sub> tract explain pure alleles as short as (CGG)<sub>8</sub> and as long as (CGG)<sub>31</sub> (Figure 2a). Interestingly, we observed a fully expanded haplotype (A<sub>3</sub>) deriving a single DXS<sub>998</sub> mutation from A<sub>1</sub> found in normal pure alleles. Both hypotheses of (1) a de novo expansion from (CGG)<sub>24</sub> and (2) recombination or two mutations on DXS<sub>998</sub> can be raised, although the second is more likely. On the other hand, to test the hypothesis of loss of AGG interruptions, we looked at genetic distances between pure and interrupted normal alleles, and observed three shared haplotypes (A<sub>1</sub>, A<sub>2</sub> and A<sub>4</sub>) in the main torso of the network (Figure 2b). The only haplotype found exclusively in a pure normal allele (A<sub>13</sub>-(CGG)<sub>31</sub>) derives a single FRAXAC<sub>2</sub> mutation from A<sub>4</sub>. Taking into account that an allele with 31 repeats (CGG)<sub>10</sub>AGG(CGG)<sub>20</sub> has also been found to segregate on the A<sub>4</sub> haplotype, the loss of the AGG interruption (through a recurrent mutation in the first base: AGG<sub>4</sub>CGG) is also possible to be on the origin of this (CGG)<sub>31</sub> pure allele. To explain the ancestry of the (CGG)<sub>8</sub> allele in our analyzed 123 normal A-T-T alleles, a multistep mutation involving the gain of several CGG units must have occurred as no intermediate sized alleles from lineage A have been found between 8 and 20 CGGs. While analyzing our four STRs in cohorts 1 and 2, we faced again the problem of FRAXAC<sub>2</sub> heterogeneity. Here, to discard the hypothesis of biased results due to uncontrolled source of variation within this marker, we reconstructed phylogenetic networks based only on DXS<sub>998</sub>, DXS<sub>548</sub> and FRAXAC<sub>1</sub> markers; results obtained led to the same conclusions (Supplementary Figure S1). Similarly, to gain insight into the noise introduced in our networks due to recombination, we excluded the most distant marker (DXS<sub>998</sub> located several blocks apart from the repeat); again,



we obtained similar results suggesting that recombination was not a major source of additional variation to analyzed STR haplotypes (Supplementary Figure S2).

## Discussion

In fragile X syndrome, the occurrence of mosaic full mutations, that is, multiple alleles within the expanded range, has been fairly reported.<sup>26–28</sup> This is owing to the high intergenerational and somatic instability of fully expanded alleles. Rarer is the presence of both premutation and full mutation in the same individual, which usually underlies the origin of de novo expansions.<sup>10,29</sup> In this study, we presented four cases of an even rarer type of intra-tissue size mosaicism, with both normal and fully mutated alleles in different blood cells from a single male patient. The presence of normal/intermediate FMR1 alleles in addition to the full mutation did not modify the typical FXS clinical presentation observed in all the four patients. This may be due to the low levels of the short alleles when compared with the expanded ones and/or to their methylation status as all mosaic patients displayed methylated full expansions and unmethylated normal and intermediate alleles. This differed from the mosaic case presented by Basuta et al., a male patient with FXTAS, who showed a completely unmethylated FMR1 full mutation in peripheral blood mononucleated cells (as well as intra- and intertissue size mosaicism, with premutation alleles of different sizes in fibroblasts and sperm).<sup>30</sup> In this case, the complex phenotype (features of FXS and late onset neurological deterioration with probable FXTAS) was explained by a combined molecular mechanism, derived from both the elevated FMR1 mRNA levels and low FMRP expression. In our mosaic patients, after genotyping the respective mothers for FMR1, we have shown that all the four large contractions occurred independently in an early embryonic development of the mosaic patients, from a fully expanded to a normal-sized (CGG)<sub>n</sub> allele. Therefore, we benefited from having both parental and derived alleles to unravel the mechanisms of FMR1 CGG repeat instability, more specifically of large contractions. It would be interesting to follow the repeat dynamics of our mosaic cases, namely by analyzing the sperm cells of these patients. Interestingly, a bias for repeat contraction has been suggested in males since a man carrying an unmethylated FMR1 full mutation in peripheral blood mononucleated cells presented a premutation allele with 86 CGG repeats in sperm cells.<sup>30</sup> Taking into account that maternal transmissions result almost invariably in expansions of the CGG allele size, one could hypothesize the maintenance of highly unstable pure CGGs owing to the existence of expansion-prone alleles in female meioses, which would contract again to the normal range in the paternal lineage during father–offspring transmissions. Despite several studies have included both SNP and STR markers flanking the repetitive region, herein, to follow more accurately the recent evolution of the FMR1 repeat, we applied a population genetics approach by first analyzing stable SNP backgrounds to next place fast-evolving STRs within the pre-identified SNP lineages. This is, to our knowledge, the first study combining such a structured haplotype analysis with seven flanking polymorphic markers phased with FMR1 alleles (with known repeat size and configuration). A single SNP background was not shown to segregate with (CGG)<sub>n</sub> alleles from our analyzed mosaics, however, the shared extended DXS<sub>998</sub>-DXS<sub>548</sub>-FRAXAC<sub>1</sub>-FRAXAC<sub>2</sub> haplotype within each lineage (34-44-38-336 in lineage A and the 3′-end 36-333 in lineage C) raised the question of a lineage-specific risk haplotype more prone to large contractions. After observing that all contracted

alleles displayed no AGG interruptions, we searched for the answer by analyzing STR haplotypes of normal pure FMR1 alleles. The close phylogenetic relationships observed among normal pure repeats from the control population suggested their recent evolution from a common ancestor instead of recurrent large contractions from the pool of expanded alleles. Nevertheless, even if drastic reductions on FMR1 repeat sizes are not on the origin of all pure normal alleles, the occurrence of these rare events on recurrent extended STR-based haplotypes may indicate that a predisposing factor to large contractions may be tagged in both lineages A and C. The mechanisms behind this cis-effect may be via (1) changes in chromatin organization as SNP variants may impede the binding of CTCF, an insulator DNA-binding protein previously linked to repeat instability;<sup>31</sup> (2) gene regulation if these or associated SNPs alter the binding of transcription factors as transcription through the repeat tract has been shown to increase repeat instability;<sup>32</sup> and/or (3) differential interaction of repair proteins with the SNP variants at FMR1, resulting in different instability patterns as suggested for other trinucleotide repeat disorders.<sup>33</sup>

In the literature, a few other multistep contractions have been described in the maternal lineage (although in a smaller range), namely the change of interrupted premutated alleles either to smaller premutations or to intermediate pure alleles.<sup>10</sup> It would be interesting to assess the haplotype background where these events took place and gain insight whether small and large contractions share the same risk factors. While analyzing the evolution of normal pure alleles, other, less drastic multistep mutations seemed to occur. Indeed, a multistep mechanism involving the gain or loss of several repetitive units during the intergenerational transmission of expanded alleles has been suggested to explain the evolution of other disease-expansion repeats.<sup>34–36</sup>

Further studies of repeat instability underlying the different groups of normal pure (CGG)<sub>n</sub> alleles (the unstable ones resulted from large contractions and those, probably more stable, evolved through interrupted repeats) would be of utmost importance for the research of both the mechanistic aspects of repeat instability and the clinical risk-assessment for fragile X syndrome.

## Acknowledgements

We acknowledge all patients and their families for participating in this study, and medical geneticists Ana Maria Fortuna, Mariana Soeiro e Sá, Oana Moldovan and Ana Berta Sousa for referring the patients. UMIB is funded by National Funds through FCT Foundation for Science and Technology, under the Pest-OE/SAU/UI0215/2014. SM is funded by the project IF/00930/2013/CP1184/CT0002 from FCT. IM received a research grant 2014 DEFI-CHP, E.P.E.

## REFERENCES

- 1 Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A. et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–914 (1991).
- 2 Nolin, S. L., Lewis, F. A. 3rd, Ye, L. L., Houck, G. E. Jr, Glicksman, A. E., Limprasert, P. et al. Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 59, 1252–1261 (1996).
- 3 Nolin, S. L., Glicksman, A., Ding, X., Ersalesi, N., Brown, W. T., Sherman, S. L. et al. Fragile X analysis of 1112 prenatal samples from 1991 to 2010. *Prenat. Diagn.* 31, 925–931 (2011).



- 4 Nolin, S. L., Brown, W. T., Glicksman, A., Houck, G. E. Jr, Gargano, A. D., Sullivan, A. et al. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet.* 72, 454–464 (2003).
- 5 Morton, N. E. & Macpherson, J. N. Population genetics of the fragile-X syndrome: multiallelic model for the FMR1 locus. *Proc. Natl Acad. Sci. USA* 89, 4215–4217 (1992).
- 6 Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.* 31, 37–46 (2002).
- 7 Mirkin, E. V. & Mirkin, S. M. To switch or not to switch: at the origin of repeat expansion disease. *Mol. Cell* 53, 1–3 (2014).
- 8 Peixoto, A., dos Santos, M. R., Seruca, R., Amorim, A. & Castedo, S. Analysis of FMR1 and flanking microsatellite markers in normal and fragile X chromosomes in Portugal: evidence for a 'protector' haplotype. *Eur. J. Hum. Genet.* 6, 518–522 (1998).
- 9 Ennis, S., Murray, A., Brightwell, G., Morton, N. E. & Jacobs, P. A. Closely linked cis-acting modifier of expansion of the CGG repeat in high risk FMR1 haplotypes. *Hum. Mutat.* 28, 1216–1224 (2007).
- 10 Nolin, S. L., Glicksman, A., Ersalesi, N., Dobkin, C., Brown, W. T., Cao, R. et al. Fragile X full mutation expansions are inhibited by one or more AGG interruptions in premutation carriers. *Genet. Med.* 17, 358–364 (2015).
- 11 Yrigollen, C. M., Durbin-Johnson, B., Gane, L., Nelson, D. L., Hagerman, R., Hagerman, P. J. et al. AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome. *Genet. Med.* 14, 729–736 (2012).
- 12 Fernandez-Carvajal, I., Lopez Posadas, B., Pan, R., Raske, C., Hagerman, P. J. & Tassone, F. Expansion of an FMR1 grey-zone allele to a full mutation in two generations. *J. Mol. Diagn.* 11, 306–310 (2009).
- 13 Yrigollen, C. M., Sweha, S., Durbin-Johnson, B., Zhou, L., Berry-Kravis, E., Fernandez-Carvajal, I. et al. Distribution of AGG interruption patterns within nine world populations. *Intractable Rare Dis. Res.* 3, 153–161 (2014).
- 14 Ashley-Koch, A. E., Robinson, H., Glicksman, A. E., Nolin, S. L., Schwartz, C. E., Brown, W. T. et al. Examination of factors associated with instability of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 63, 776–785 (1998).
- 15 Chiurazzi, P., Macpherson, J., Sherman, S. & Neri, G. Significance of linkage disequilibrium between the fragile X locus and its flanking markers. *Am. J. Med. Genet.* 64, 203–208 (1996).
- 16 Nolin, S. L., Sah, S., Glicksman, A., Sherman, S. L., Allen, E., Berry-Kravis, E. et al. Fragile X AGG analysis provides new risk predictions for 45–69 repeat alleles. *Am. J. Med. Genet. A* 161, 771–778 (2013).
- 17 Kunst, C. B. & Warren, S. T. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 77, 853–861 (1994).
- 18 Limprasert, P., Thanakitgosate, J., Jaruthamsophon, K. & Sripo, T. Unique AGG interruption in the CGG repeats of the FMR1 gene exclusively found in Asians linked to a specific SNP haplotype. *Genet. Res. Int.* 2016, 8319287 (2016).
- 19 Gunter, C., Paradee, W., Crawford, D. C., Meadows, K. A., Newman, J., Kunst, C. B. et al. Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum. Mol. Genet.* 7, 1935–1946 (1998).
- 20 Barasoain, M., Barrenetxea, G., Ortiz-Lastra, E., González, J., Huerta, I., Télez, M. et al. Single nucleotide polymorphism and FMR1 CGG repeat instability in two Basque valleys. *Ann. Hum. Genet.* 76, 110–120 (2012).
- 21 Crawford, D. C., Schwartz, C. E., Meadows, K. L., Newman, J. L., Taft, L. F., Gunter, C. et al. Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and single-nucleotide-polymorphism haplotypes in an African American population. *Am. J. Hum. Genet.* 66, 480–493 (2000).

- 22 Pretto, D., Yrigollen, C. M., Tang, H. T., Williamson, J., Espinal, G., Iwahashi, C. K. et al. Clinical and molecular implications of mosaicism in FMR1 full mutations. *Front. Genet.* 5, 318 (2014).
- 23 Tabolacci, E., Pomponi, M. G., Pietrobono, R., Chiurazzi, P. & Neri, G. A unique case of reversion to normal size of a maternal premutation FMR1 allele in a normal boy. *Eur. J. Hum. Genet.* 16, 209–214 (2008).
- 24 Jorge, P., Oliveira, B., Marques, I. & Santos, R. Development and validation of a multiplex-PCR assay for X-linked intellectual disability. *BMC Med. Genet.* 14, 80 (2013).
- 25 Chiurazzi, P., Pomponi, M. G., Sharrock, A., Macpherson, J., Lormeau, S., Morel, M. L. et al. DNA panel for interlaboratory standardization of haplotype studies on the fragile X syndrome and proposal for a new allele nomenclature. *Am. J. Med. Genet.* 83, 347–349 (1999).
- 26 Ferreira, S. I., Pires, L. M., Ferrao, J., Sa, J., Serra, A. & Carreira, I. M. Mosaicism for FMR1 gene full mutation and intermediate allele in a female foetus: a postzygotic retraction event. *Gene* 527, 421–425 (2013).
- 27 Grasso, M., Faravelli, F., Lo Nigro, C., Chiurazzi, P., Sperandio, M. P., Argusti, A. et al. Mosaicism for the full mutation and a microdeletion involving the CGG repeat and flanking sequences in the FMR1 gene in eight fragile X patients. *Am. J. Med. Genet.* 85, 311–316 (1999).
- 28 Rife, M., Badenas, C., Quinto, L., Puigoriol, E., Tazón, B., Rodriguez-Revenga, L. et al. Analysis of CGG variation through 642 meioses in Fragile X families. *Mol. Hum. Reprod.* 10, 773–776 (2004).
- 29 Dombrowski, C., Levesque, S., Morel, M. L., Rouillard, P., Morgan, K. & Rousseau, F. Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* 11, 371–378 (2002).
- 30 Basuta, K., Schneider, A., Gane, L., Polussa, J., Woodruff, B., Pretto, D. et al. High functioning male with fragile X syndrome and fragile X-associated tremor/ataxia syndrome. *Am. J. Med. Genet.* 167A, 2154–2161 (2015).
- 31 Libby, R. T., Hagerman, K. A., Pineda, V. V., Lau, R., Cho, D. H., Baccam, S. L. et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet.* 4, e1000257 (2008).
- 32 Filippova, G. N., Thienes, C. P., Penn, B. H., Cho, D. H., Hu, Y. J., Moore, J. M. et al. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.* 28, 335–343 (2001).
- 33 Martins, S., Pearson, C. E., Coutinho, P., Provost, S., Amorim, A., Dubé, M. P. et al. Modifiers of (CAG)(n) instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum. Genet.* 133, 1311–1318 (2014).
- 34 Martins, S., Calafell, F., Wong, V. C., Sequeiros, J. & Amorim, A. A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. *Eur. J. Hum. Genet.* 14, 932–940 (2006).
- 35 Santos, D., Pimenta, J., Wong, V. C., Amorim, A. & Martins, S. Diversity in the androgen receptor CAG repeat has been shaped by a multistep mutational mechanism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 165, 581–586 (2014).
- 36 Trang, H., Stanley, S. Y., Thorner, P., Faghfoury, H., Schulze, A., Hawkins, C. et al. Massive CAG repeat expansion and somatic instability in maternally transmitted infantile spinocerebellar ataxia type 7. *JAMA Neurol.* 72, 219–223 (2015).

## TABLES

Table 1 Summary of clinical and genotyping data of mosaic male FXS patients and respective mothers

Age at					Maternal FMR1 alleles		
Table 2 SNP-STR haplotypes of normal pure FMR1 alleles observed in mosaic patients and in our cohort 1 (212 males with a FMR1 allele within a CGG normal-intermediate range)					Normal	Premutation/ Full mutation	Familial information
FMR1 alleles	FMR1 repeat	SNP lineage	STR haplotype	Haplotype ID	son for referral	Normal FMR1 mosaic allele <sup>a</sup>	Expanded FMR1 mosaic allele <sup>a</sup>
Normal alleles from mosaic cases	(CGG) <sub>35</sub>	A-T-T	34-44-38-336	A21	ily history of FXS, ich delay	UnMet (CGG) <sub>35</sub>	Met (CGG) <sub>&gt;200</sub>
	(CGG) <sub>26</sub>	A-T-T	34-44-38-336	A21	ily history of FXS, sm spectrum	UnMet (CGG) <sub>39</sub>	Met (CGG) <sub>&gt;200</sub>
	(CGG) <sub>39</sub>	G-C-C	32-42-36-333	C4	ily history of FXS, rder, speech delay	UnMet (CGG) <sub>18</sub>	UnMet (CGG) <sub>125</sub> and Met (CGG) <sub>&gt;200</sub>
	(CGG) <sub>18</sub>	G-C-C	36-52-36-333	C25 <sup>a</sup>			
Normal pure alleles from cohort 1	(CGG) <sub>8</sub>	A-T-T	34-38-38-336	A4	ile X syndrome.		
	(CGG) <sub>17</sub>	G-C-C	32-40-36-333	C9	n obtained by Southern blot after methylation sensitive endonuclease digestion.		
	(CGG) <sub>23</sub>	A-T-T	34-38-38-336	A4	s of mosaic alleles is denoted by Met (methylated) and UnMet (unmethylated).		
	(CGG) <sub>24</sub>	A-T-T	34-40-38-336	A1			
	(CGG) <sub>24</sub>	A-T-T	36-40-38-336	A2			
	(CGG) <sub>26</sub>	G-T-C	32-48-36-333	B18			
	(CGG) <sub>31</sub>	A-T-T	34-38-38-338	A13			
	(CGG) <sub>31</sub>	G-T-C	36-52-42-337	B15			

Abbreviations: SNP, single-nucleotide polymorphism; STR, short tandem repeat.  
<sup>a</sup>Haplotype not found in cohorts 1 or 2.

Table 3 Relative frequencies of FMR1 lineages genotyped in the four mosaic FXS patients in comparison with those observed in normal pure, normal interrupted and expanded (CGG)<sub>n</sub> alleles

FMR1 lineage	Mosaics (n = 4)	Normal alleles		
		Normal pure alleles (n = 8)	with one AGG interruption (n = 55)	Expanded alleles (n = 123)
A (A-T-T)	50%	62.5%	60%	10.6%
B (G-T-C)	—	25%	25.5%	59.3%
C (G-C-C)	50%	12.5%	7.3%	22.8%
D (G-T-T)	—	—	1.8%	2.4%
E (A-T-C)	—	—	5.4%	4.9%

Abbreviation: FXS, Fragile X syndrome.

## FIGURE

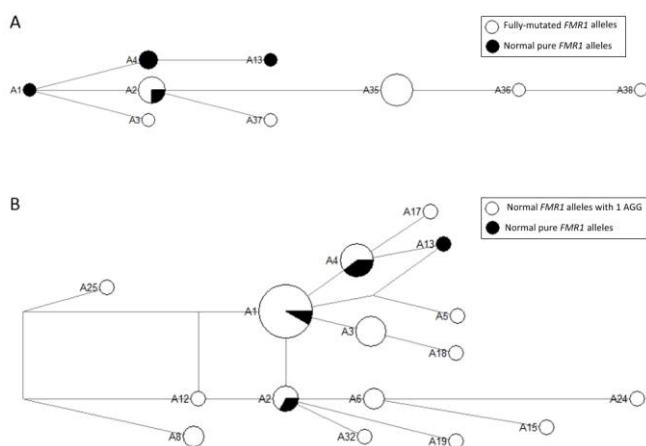


Figure 1 Phylogenetic networks showing the most parsimonious relationships among FMR1 haplotypes (DXS998-DXS548-FRAXAC1-FRAXAC2) from lineage A (A-T-T) to test the hypothesis of normal pure CGGs originated by (a) the contraction of fully expanded alleles and (b) the loss of AGG interruptions. Circle and line sizes are proportional to the number of individuals and stepwise mutations, respectively. Squares indicate recombinations.

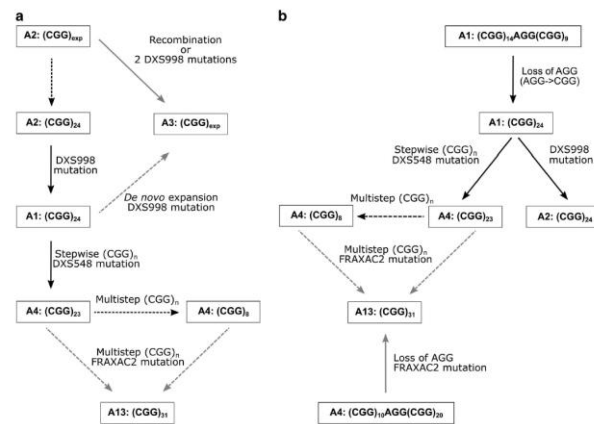


Figure 2 Schematic representation of the two hypotheses for the origin of pure normal FMR<sub>1</sub> alleles in the control population by (a) large contractions from the pool of fully expanded alleles and (b) the loss of AGGs from the pool of normal interrupted alleles. Dashed lines denote multistep mutations at FMR<sub>1</sub>; gray lines indicate alternative hypotheses.